5-5-2022

Reinforcement Learning Algorithms and Complexity of Inventory Control, A Review

Aravindh Sekar
Dakota State University, aravindh.sekar@trojans.dsu.edu

David Zeng
Dakota State University, david.zeng@dsu.edu

Reinforcement Learning Algorithms and Complexity of Inventory Control – A Review

Aravindh Sekar
Dakota State University
Aravindh.Sekar@trojans.dsu.edu

David Zeng
Dakota State University
David.Zeng@dsu.edu

ABSTRACT

Driven by the ability to perform sequential decision-making in complex dynamic situations, Reinforcement Learning (RL) has quickly become a promising avenue to solve inventory control (IC) problems. The objective of this paper is to provide a comprehensive overview of the IC problems that have been effectively solved due to the application of RL. Our contributions include providing the first systematic review in this field of interest and application. We also identify potential extensions and come up with four propositions that formulate a theoretical framework that may help develop RL algorithms to solve complex IC problems. We recommend specific future research directions and novel approaches to solving IC problems.

Keywords
Reinforcement Learning, Inventory Control, literature review, propositions.

INTRODUCTION

RL framework can be simple and flexible and thereby making it possible to apply to many different problems in many ways. RL focuses on sequential decision-making problems and deals with learning via interaction and feedback (Boute et al., 2021). This paper aims at delivering an objective overview of the developments that have taken place in the Inventory Control section using Reinforcement learning. We summarize our intentions by proposing the following research questions: 1) To what extent is Reinforcement Learning utilized in Inventory Control? 2)What limitations does the current application of RL have in inventory Control? 3) What theoretical component is generated that when used, may help resolve potential complex IC problems?

Our analysis indicates that Hybrid algorithms may provide more stability and may help in fixing more unsolved IC problems. For a specific replenishment inventory, we realize that off-policy algorithms, that learn from existing data, work best in identifying the right policy.

The contribution of this research is three-fold. 1) To our knowledge, this is the first to provide a systematic review on the application of Reinforcement learning in inventory control problems. 2) We identify and quantify the extent of the application of Reinforcement learning and its algorithms in particular sections of inventory control. 3) We recognize the potential extensions of the current application and come up with four sets of propositions thereby shedding light on the potential opportunities for future researchers and even for organizations as the first step to guide future research. The paper is organized as follows. After summarizing the research methodology, we analyze the related work to connect the types of RL algorithms and the characteristics of the specific IC problems. Our review produces four general observations (propositions) and a two-dimensional framework that can be used to identify research gaps. We conclude by discussing future research opportunities.

SYSTEMATIC REVIEW METHODOLOGY

For the selection and screening of research articles for literature review, we referred to the PRISMA model primarily developed for reporting systematic reviews (Liberati et al., 2009). The model is bifurcated into four steps 1) identification, 2) screening, 3) eligibility, and 4) included.

Using the identified Keywords, three databases were searched: EBSCO host, Science Direct, and IEEE. After identification of papers, we applied four screening criteria: 1) Only articles published between 2008 and 2021, 2) Only full published English Articles and 3) Only Journal Articles. After screening, we applied exclusion criteria by 1) removing duplicates, and 2) relevance of title and abstract based on the research. One article by Giannoccaro and Pontrandolfo (2002) was taken outside the year range due to it being one of the earlier papers talking about RL application in IC. The final number of articles was 15. The final filter applied was based on going through the entire article and making sure it aligned with our research objective.

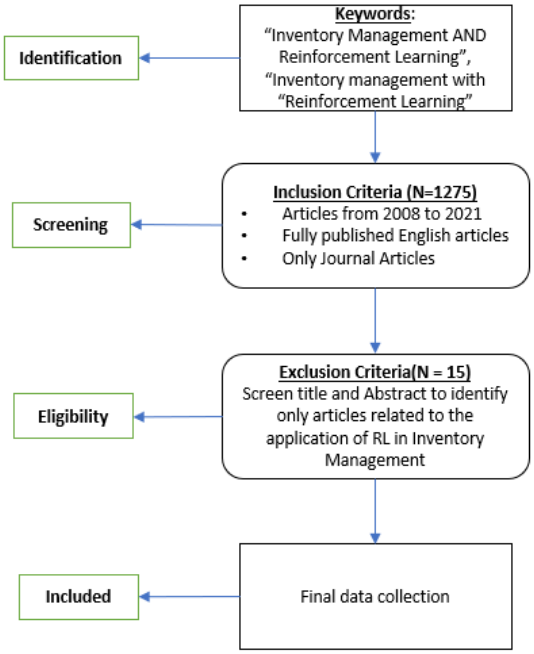

Figure 1. Systematic Review Methodology

ADDRESSING THE RESEARCH GAPS

In Inventory Control, RL is used in several sections such as determining ordering policy considering only one company in every stage (Giannoccaro and Pontrandolfo, 2002) and in optimizing replenishment quantity from the multi-stages supply chain, using a single actor and under deterministic demand and lead time (Chaharsooghi et al., 2008; Mortazavi et al., 2015). By identifying gaps in the literature, we aim to address the research questions stated in the Introduction.

In the literature, a variety of RL algorithms have been applied in IC situations. Q-learning, an RL method, is used in formulating ordering and pricing problems in a multi-retailer environment (Dogan and Güner, 2015). Q-learning was also used to frame an ordering policy with uncertain lead times and uncertain customer demand (Chaharsooghi et al., 2008). Q-learning with newsvendor rule with a vendor-managed inventory system is used to resolve replenishment problems where the supplier decides the needed inventory for retailers (Sui et al., 2010). Q-learning, using stochastic demand and a fixed lead time, formulated a perishable specific inventory policy (Kara and Dogan, 2018).

Advantage Actor-Critic (A3C), a combination of policy and value iteration, is used in solving three classic inventory problems, Lost sales, dual sourcing, and multi-echelon inventory management. Another hybrid algorithm, Proximal Policy Optimization (PPO), defined retail inventory control as sourcing raw materials for synchronization and business continuity (Kegenbekov and Jackson, 2021). A comprehensive overview of the identified papers that are categorized by RL design choices is provided in Table 1.

Author	Algorithm	Value Iteration	Hybrid Iteration	Off Policy	On Policy	Targeted Inventory Issues
Gijsbrechts et al., 2019	A3C		✓		✓	Lost Sales, Dual Sourcing, and Multi-echelon Inventory Management
Sui et al., 2010	Q-learning with newsvendor Rule	✓		✓		Supplier based Replenishment policy
Zarandi et al., 2013	FFRL	✓		✓		Inventory Replenishment based on retailer value.
Wang et al., 2021	PAQ-DQN/PAQ-A2C	✓	✓	✓	✓	Pricing and Inventory Replenishment for Maximum Profits
Kara and Dogan, 2018	Q-learning and Sarsa	✓		✓		Age-based Inventory Replenishment policy
Giannoccaro and Pontrandolfo, 2002	SMPD process with a SMART algorithm	✓		✓		Optimized Inventory Replenishment policy
Kegenbekov and Jackson, 2021	PPO		✓		✓	Synchronized Inventory Replenishment
Vanvuchelen et al., 2020	PPO		✓		✓	Joint Replenishment Policy focusing on multiple items
Jiang and Sheng, 2009	Case-Based RL	✓		✓		Dynamic Inventory Control in a multi-agent setting
Kwon et al., 2008	Case-Based Myopic RL	✓		✓		Inventory Replenishment with time-dependent inventory and large state spaces
Sun et al., 2019	Q-learning and DQN	✓		✓		Inventory Replenishment for fresh products.
Dogan and Güner, 2015	Q-learning	✓		✓		Inventory Replenishment with pricing
Oroojlooy Jadid et al., 2017	A DQN Variant Algorithm	✓		✓		Inventory Replenishment
Chaharsooghi et al., 2008	Q learning-based - RL Ordering Mechanism	✓		✓		Inventory Replenishment
Zhou and Zhou, 2019	DQN + Improvement Heuristics (IH)	✓		✓		Optimal policy for a multi-echelon inventory

Table 1. Summary of Reinforcement Learning Applications in Inventory Control

Figure 2 shows an overview of classes of RL algorithms that takes off from the identified literature and shows the different sections of IC that RL has been applied. The categorization shows that Hybrid algorithms are used for solving specific inventory control issues such as Lost Sales and Dual Sourcing. But some require extensive hyperparameter tuning resulting in the emergence of PPO. Figure 2 also shows how off-policy algorithms are extensively used in Inventory Replenishment.

Inventory policies have been defined using simple single-agent models (Giannoccaro and Pontrandolfo, 2002; Wang et al., 2021). Inventory policies for perishable products were formulated using a single agent and single echelon or one wholesaler and one retailer model or a multi-retailer and a single supplier model (Kara and Dogan, 2018; Sun et al., 2019; Dogan and Güner, 2015). However, multiple agents are usually involved in supply chains and their interactions may have a big impact on how the policies are formulated. This leads us to frame four propositions that consider what has been done so far and provide a theoretical component that can be used in identifying the right RL method and ultimately make it applicable for different IC scenarios.

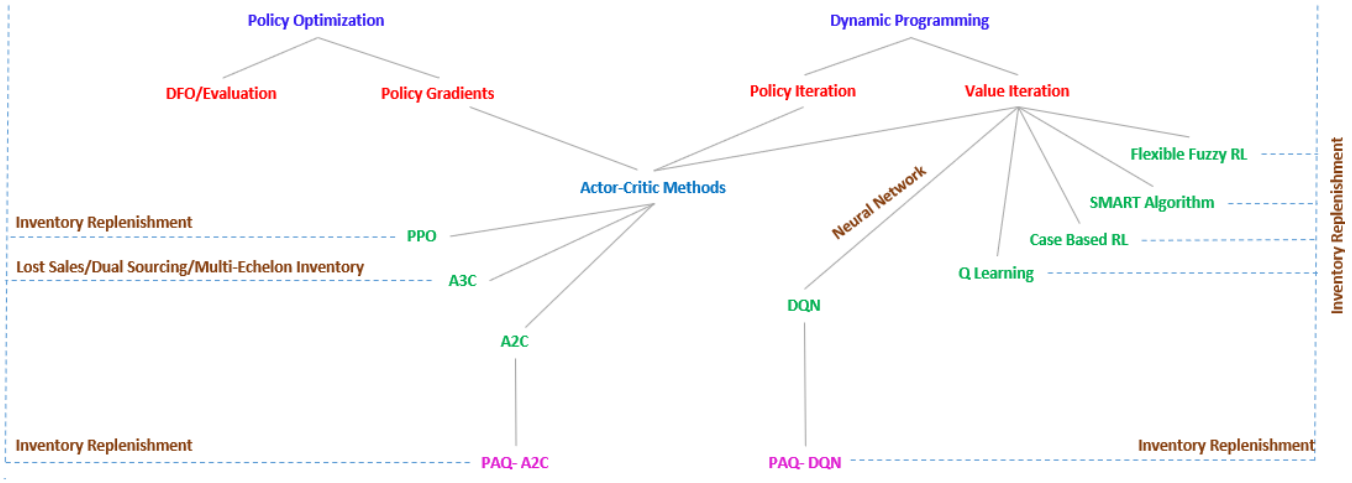

Figure 2. Reinforcement Learning Algorithms and its application in Inventory Management

When the state-action space is large, exploration in RL becomes difficult and this is called the curse of dimensionality. Dynamic Programming suffers from this curse making them computationally intractable for sizeable problems. This can be overcome by approximating the Q-values in the Q-learning algorithm with a deep neural network (Sutton and Barto, 1998). This leads to the proposition1 below:

Proposition1: Applying RL with a Deep neural network helps overcome the curse of dimensionality in inventory problems.

DQN algorithm which uses a deep neural network to obtain an approximation of the Q-function overcomes the curse of dimensionality and trains it through the iterations of the Q-learning algorithm while updating another target network (Mnih et al., 2015). The curse of dimension can also be reduced by applying approximate dynamic programming (Gijsbrechts et al., 2019). In short, Deep RL provides a way to avoid this curse of dimensionality (Vanvuchelen et al., 2020).

For organizations, it is challenging to handle perishable products (Karaesmen et al., 2011). It is important to know the customer demand, time on the shelf, and the necessary lead time situation before formulating an inventory policy. This challenge can be handled with access to historical data and off-policy algorithms can use existing data and help in learning robust policies.

Proposition2: Off-policy RL algorithms are mostly used in determining near-optimal replenishment policy for perishable products.

Deep Q-network was used to investigate a joint pricing and inventory control problem and obtain near-optimal pricing and replenishment policies for perishable inventory systems with positive lead time (Wang et al., 2021). An off-policy algorithm, Q-learning, was used to study dynamic inventory control issues for perishable products with positive lead time and fixed timelines (Kara and Dogan, 2018). Off policy RL approach was followed to manage inventory decisions at all stages of a synchronized supply chain thereby optimizing the performance of the whole supply chain (Giannoccaro and Pontrandolfo, 2002).

When formulating replenishment policies, organizations need to consider the multiple layers (echelons) of distribution centers and suppliers, multiple products, multiple periods, and multiple uncertainties related to product demand and prices to avoid any unexpected occurrences in their supply chain (Leonard, 2021). We expect that optimal replenishment policies can be learned with RL algorithms.

Proposition3: An optimal RL inventory replenishment policy is formulated when it considers multiple agents (suppler/retailers/customers) and multiple echelons (factories, wholesalers, distribution centers)

A replenishment policy targeting satisfying service levels based on a multi-agent system but with a single supplier was formulated (Jiang and Sheng, 2009). An inventory replenishment policy based on simultaneous ordering and pricing decisions for retailers working in a multi-retailer competitive environment, and a near-optimal replenishment policy with stochastic demand and a fixed lead time were formulated using a single agent as the base (Dogan and Güner, 2015; Kara and Dogan, 2018).

Value-based methods use neural networks to approximate the optimal action-value functions and are most fruitful when data is scarce or gathering new data is difficult. Policy-based methods directly approximate the optimal inventory policy by minimizing the policy loss in each update of the neural network (Boute et al., 2021). A combination of both value and policy iteration is called the actor-critic method or hybrid approach. Leveraging both value and Policy iterations provides more stability and adds a value-based baseline to policy-based methods (Mnih et al., 2015). Thus, we propose that these algorithms help organizations in solving IC problems.

Proposition4: RL algorithms with hybrid iterations solve intractable inventory control issues compared to non-hybrid algorithms.

A3C algorithm was used to solve intractable inventory problems such as dual sourcing, lost sales, and multi-echelon inventory problems (Gijsbrechts et al., 2019). But A3C requires extensive hyperparameter tuning which is computationally costly. This was overcome by using PPO, due to its ability to converge properties while being more sample efficient and easier to implement (Vanvuchelen et al., 2020). PPO was used to demonstrate a supply chain policy by synchronizing inbound and outbound flows, in a supply chain leading to end-to-end visibility (Kegenbekov and Jackson, 2021).

In Figure 3, we associate papers in each quadrant and identify associations between the propositions based on key design categories. two design categories have been identified - Off/On-Policy and Value /Policy-Based. On-policy algorithms work with a single policy and require any observations to have been generated using that policy (Hausknecht and Stone, 2016). Off-policy models can use experience from other policies, such as tuples generated from older versions of the neural network or tuples from a data set that was collected upfront, which it stores in an experienced buffer (Boute et al., 2021).

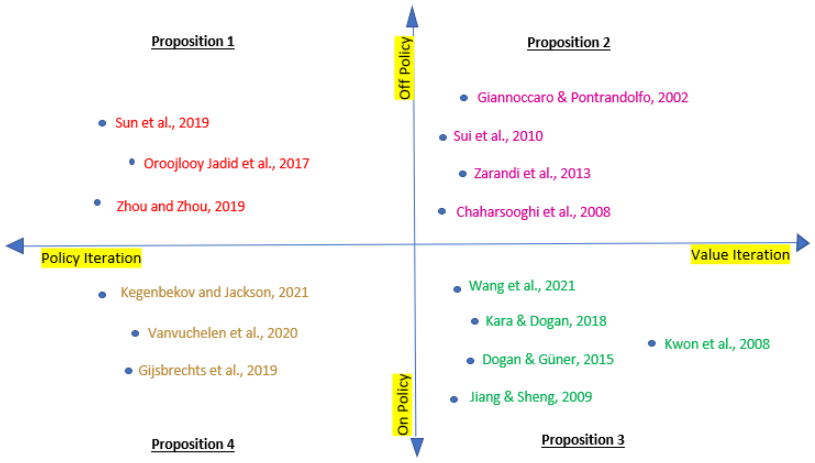

Figure 3. Proposition Quadrant and its corresponding papers

The identified propositions are interconnected based on the two design categories. P1 and P2 connect based on the application of neural networks to approximate the action-value functions. P3 and P4 go off the combination of both Value and policy-based methods because of the need to find an optimal policy for complex IC scenarios and the associated computational costs.

CONCLUSION

Based on the connections between the types of RL algorithms and the nature of IC problems they aim to solve, we formulated four propositions indicating that Hybrid algorithms when used in complex unknown situations, provide more stability in resolving IC problems. We also identify that off-policy value-based algorithms play a major role in determining the right policy for inventory replenishment. These propositions can shed light on the potential opportunities for developing novel approaches to solving inventory control problems.

Future research may explore the application of Hybrid algorithms, specifically PPO, to solve specific inventory control problems such as Overstocking, Dual Sourcing, and Lost Sales. Further research can be performed by accessing historical data, lead times, and previous challenges thereby utilizing off-policy value-based algorithms in identifying the right optimal policy for organizations to follow (Boute et al., 2021). Additionally, mapping Model-Free/Model-based algorithms with characteristics of specific IC problems may be promising. Testing the propositions with historical data and in simulated environments may pave a way for novel RL applications in complex IC situations.

REFERENCES

1. Boute, R., Gijsbrechts, J., Jaarsveld, W., and Vanvuchelen, N. (2021). Deep Reinforcement Learning for Inventory Control: A Roadmap. European Journal of Operational Research. https://doi.org/10.1016/j.ejor.2021.07.016
2. Chaharsooghi, S. K., Heydari, J., and Zegordi, S. H. (2008). A reinforcement learning model for supply chain ordering management: An application to the beer game. Decision Support Systems, 45(4), 949–959. https://doi.org/10.1016/j.dss.2008.03.007
3. Dogan, I., and Güner, A. R. (2015). A reinforcement learning approach to competitive ordering and pricing problem. Expert Systems, 32(1), 39–48. https://doi.org/10.1111/exsy.12054
4. Giannoccaro, I., and Pontrandolfo, P. (2002). Inventory management in supply chains: A reinforcement learning approach. International Journal of Production Economics, 78(2), 153–161. https://doi.org/10.1016/S0925-5273(00)00156-0
5. Gijsbrechts, J., Boute, R., Zhang, D., and Van Mieghem, J. (2019). Can Deep Reinforcement Learning Improve Inventory Management? Performance on Dual Sourcing, Lost Sales, and Multi-Echelon Problems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3302881
6. Hausknecht, M., and Stone, P. (2016). On-Policy vs. Off-Policy Updates for Deep Reinforcement Learning. 7.
7. Jiang, C., and Sheng, Z. (2009). Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. Expert Systems with Applications, 36(3), 6520–6526. https://doi.org/10.1016/j.eswa.2008.07.036
8. Kara, A., and Dogan, I. (2018). Reinforcement learning approaches for specifying ordering policies of perishable inventory systems. Expert Systems with Applications, 91, 150–158. https://doi.org/10.1016/j.eswa.2017.08.046
9. Karaesmen, I., Scheller-wolf, A., and Deniz, B. (2011). Managing Perishable and Aging Inventories: Review and Future Research Directions. In Planning Production and Inventories in the Extended Enterprise (pp. 393–436). https://doi.org/10.1007/978-1-4419-6485-4_15
10. Kegenbekov, Z., and Jackson, I. (2021). Adaptive Supply Chain: Demand-Supply Synchronization Using Deep Reinforcement Learning. Algorithms, 14(8), 240. http://www.ezproxy.dsu.edu:2087/10.3390/a14080240
11. Kwon, I., Kim, C., Jun, J., and Lee, J. (2008). Case-based myopic reinforcement learning for satisfying target service level in the supply chain. Expert Systems with Applications, 35(1–2), 389–397. https://doi.org/10.1016/j.eswa.2007.07.002
12. Leonard, M. (2021, November 15). How the pandemic has affected the just-in-time inventory approach. Construction Dive. https://www.constructiondive.com/news/inventory-lean-just-in-time-shortage-supply-chain/610029/
13. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., and Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Medicine, 6(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

14. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236
15. Mortazavi, A., Arshad khamseh, A., and Azimi, P. (2015). Designing of an intelligent self-adaptive model for supply chain ordering management system. Engineering Applications of Artificial Intelligence, 37, 207–220. https://doi.org/10.1016/j.engappai.2014.09.004
16. Oroojlooy jadid, A., Nazari, M., Snyder, L., and Takáč, M. (2017). A Deep Q-Network for the Beer Game: A Reinforcement Learning Algorithm to Solve Inventory Optimization Problems. Neural Information Processing Systems (NIPS), Deep Reinforcement Learning Symposium 2017.
17. Sui, Z., Gosavi, A., and Lin, L. (2010). A Reinforcement Learning Approach for Inventory Replenishment in Vendor-Managed Inventory Systems With Consignment Inventory. Engineering Management Journal, 22(4), 44–53. https://doi.org/10.1080/10429247.2010.11431878
18. Sun, R., Sun, P., Li, J., and Zhao, G. (2019). Inventory Cost Control Model for Fresh Product Retailers Based on DQN. 2019 IEEE International Conference on Big Data (Big Data), 5321–5325. https://doi.org/10.1109/BigData47090.2019.9006424
19. Sutton, R. S., and Barto, A. G. (2018). Reinforcement learning: An introduction (Second edition). The MIT Press.
20. Vanvuchelen, N., Gijsbrechts, J., and Boute, R. (2020). Use of Proximal Policy Optimization for the Joint Replenishment Problem. Computers in Industry, 119, 103239. https://doi.org/10.1016/j.compind.2020.103239
21. Wang, R., Gan, X., Li, Q., and Yan, X. (2021). Solving a Joint Pricing and Inventory Control Problem for Perishables via Deep Reinforcement Learning. Complexity, 2021, 1–17. https://doi.org/10.1155/2021/6643131
22. Zarandi, M., Moosavi, S., and Zarinbal, M. (2013). A fuzzy reinforcement learning algorithm for inventory control in supply chains. International Journal of Advanced Manufacturing Technology, 65(1–4), 557–569. https://doi.org/10.1007/s00170-012-4195-z
23. Zhou, J., and Zhou, X. (2019). Multi-Echelon Inventory optimizations for Divergent Networks by Combining Deep Reinforcement Learning and Heuristics Improvement. 2019 12th International Symposium on Computational Intelligence and Design (ISCID), 69–73. https://doi.org/10.1109/ISCID.2019.00023